



Constructions définitives des tables du Lexique-Grammaire : Le cas des complétives et des infinitives

Elsa Tolone

► To cite this version:

Elsa Tolone. Constructions définitives des tables du Lexique-Grammaire : Le cas des complétives et des infinitives. MajecSTIC 2010, Oct 2010, Bordeaux, France. electronic version (4 pp.). hal-00515326v2

HAL Id: hal-00515326

<https://hal.science/hal-00515326v2>

Submitted on 22 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constructions définitives des tables du Lexique-Grammaire : Le cas des complétives et des infinitives

Elsa Tolone¹

1 : Laboratoire d'informatique Gaspard-Monge, Université Paris-Est, Cité Descartes, 5 bd Descartes, Champs-sur-Marne, 77454 Marne-la-Vallée Cedex 2 - France.

Contact : elsa.tolone@univ-paris-est.fr

Résumé

Les tables du Lexique-Grammaire constituent un lexique syntaxique très riche pour le français. Cette base de données linguistique n'est cependant pas directement exploitable informatiquement car elle est incomplète et manque de cohérence. Chaque table regroupe un certain nombre d'entrées jugées similaires car elles acceptent des propriétés communes qui ne sont pas codées dans les tables mais uniquement décrites dans la littérature. Notre objectif est donc de définir pour chaque table ces propriétés indispensables à l'utilisation des tables dans les diverses applications de Traitement Automatique des Langues (TAL), telle que l'analyse syntaxique.

Abstract

Lexicon-Grammar tables are a very rich syntactic lexicon for the French language. This linguistic database is nevertheless not directly suitable for use by computer programs, as it is incomplete and lacks consistency. Tables are defined on the basis of features which are not explicitly recorded in the lexicon. These features are only described in literature. Our aim is to define for each tables these essential properties to make them usable in various Natural Language Processing (NLP) applications, such as parsing.

Mots-clés : Traitement Automatique des Langues, lexique syntaxique, Lexique-Grammaire

Keywords: Natural Language Processing, syntactic lexicon, Lexicon-Grammar

1. Introduction

Les tables du Lexique-Grammaire constituent aujourd'hui une des principales sources d'informations lexicales syntaxiques pour le français. Leur développement a été initié dès les années 1970 par Maurice Gross [2], au sein du LADL (Laboratoire d'Automatique Documentaire et Linguistique) puis du LIGM (Université Paris-Est) [3]. Ces informations se présentent sous la forme de *tables*. Chaque table correspond à une *classe* qui regroupe les éléments lexicaux d'une catégorie grammaticale donnée (verbes, noms, adjectifs, etc.) partageant certaines propriétés. Une table se présente sous forme de matrice : en lignes, les éléments de la classe correspondante ; en colonnes, les propriétés qui ne sont pas forcément respectées par tous les éléments de la classe ; à la croisée d'une ligne et d'une colonne le signe + ou - selon que l'entrée lexicale décrite par la ligne accepte ou non la propriété décrite par la colonne.

Il existe 61 tables de verbes simples, la catégorie la mieux décrite, 32 tables d'adverbes (adverbes en *-ment* et locutions adverbiales), 59 tables de noms prédicatifs (noms avec argument(s) qui sont étudiés avec leur verbe support) et 65 tables d'expressions figées. Une liste exhaustive a déjà été donnée dans [6]. Une partie est téléchargeable sous une licence libre (LGPL-LR)¹.

Chaque classe regroupe un certain nombre d'entrées jugées similaires car elles acceptent des propriétés communes, que l'on appelle les *propriétés définitives*. Elles sont en général constituées d'au

1. <http://infolingu.univ-mlv.fr>, Données Linguistiques > Lexique-Grammaire > Visualisation

moins une *construction*, dite « de base ». Par exemple, la construction N0 V à N1 indique que la phrase contient un sujet, suivi du verbe correspondant à l'entrée et d'un complément indirect introduit par la préposition à. Par ailleurs, la *propriété distributionnelle* N0 = : Nhum spécifie que le sujet doit être de type humain. Avoir les deux propriétés précédentes acceptées en même temps revient à admettre la construction N0hum V à N1.

Les constructions de base ont subi des changements au cours des années alors que ce sont les plus importantes à définir. L'objectif de cet article est de suivre leur trace pour comprendre le sens de leur modification. Ce travail a été fait pour l'ensemble des catégories [8], à savoir les verbes, les noms prédictifs, les expressions figées et les adverbes, mais nous ne détaillerons ici que les verbes, et plus particulièrement les complétives et les infinitives.

2. Les classes des verbes de M. Gross : Les complétives et les infinitives

Les classes des verbes de M. Gross sont au nombre de 19 lors de leur création [2] : de 1 à 3 les classes des infinitives et de 4 à 19 celles des complétives. Nous allons détailler pour chaque table quelle est sa construction de base actuelle et expliquer quelles sont les modifications qui ont été nécessaires.

Une seule table n'a subi aucun changement, il s'agit de la table 16 décrite dans [2] (p.208), qui a pour construction de base N0 V Prép N1 Prép N2. La table a en effet les deux prépositions et la distribution des arguments N0, N1 et N2 codées dans la table.

2.1. Modifications de la construction de base

La construction de base de la table 1 décrite dans [2] (p.160) est N0 U Prép V0-inf W. Tout d'abord, nous pouvons remarquer que la préposition peut être vide. En effet, la colonne Prép1 contenait des –, aujourd'hui remplacés par des <E>. Une proposition avait été d'avoir comme construction de base N0 U (E+Prép) V0-inf W, mais cela introduit un « OU exclusif », notation que l'on cherche à éviter pour ne pas rendre les intitulés ambigus.

Comme l'explique [2], un verbe U apparaissant dans une phrase de forme N0 U Prép V0-inf W peut ne pas imposer de contraintes sur N0. Autrement dit, étant donné une phrase quelconque N0 V0-inf W, il est possible d'y insérer un U sans que les relations entre N0 et V0-inf W en soient affectées. Considérons par exemple la phrase :

(Pierre+Cette affaire+Cette table+Faire ceci) amuse Paul,

et insérons-y le verbe *continuer* à -inf, nous obtenons la séquence :

(Pierre+Cette affaire+Cette table+Faire ceci) continue à amuser Paul,

qui est tout à fait acceptable, et la relation entre N0 et V0-inf W reste inchangée.

Or seul un petit nombre de verbes correspondent à la définition, à savoir uniquement ceux de la table 1. Cette information n'étant pas exploitable informatiquement, nous avons remplacé U par V. La construction de base de la table 1 est donc N0 V Prép V0-inf W.

La construction de base de la table 3 décrite dans [2] (p.168) est N0 V N1 V1-inf W. Il s'avère que l'on peut toujours insérer un argument locatif (Loc N2) avant le verbe à l'infinitif. Nous avons donc retenu la construction de base la plus complète, qui est N0 V N1 Loc N2 V1-inf W, ce qui n'empêche pas de garder également N0 V N1 V1-inf W comme construction de base. De plus, l'infinitive peut être supprimée pour toutes les entrées, donc N0 V N1 Loc N2 est acceptée comme troisième construction de base. Prenons un exemple illustrant ces trois constructions de base :

On a emmené Luc à l'hôpital se faire soigner

On a emmené Luc à l'hôpital

On a emmené Luc se faire soigner

2.2. Éclatement en plusieurs classes

Nous avons procédé à l'éclatement en plusieurs classes quand cela était nécessaire. C'est le cas de la table 2 dont la construction de base décrite dans [2] (p.165) est N0 V (E+N1) V0-inf avec un OU exclusif. Cela signifie que certaines entrées acceptent un argument N1 direct et d'autres non, sans que cela ne soit codé. Or, une propriété dite définitoire pour une classe est considérée comme étant acceptée pour toutes les entrées, sans exception. Nous avons donc créé une nouvelle table 2T regroupant toutes les entrées transitives (acceptant un N1) tout en les supprimant de la table 2.

Cela a été fait manuellement pour chaque entrée car aucune indication n'était donnée dans la littérature.

De plus, après avoir fait ce découpage, il s'avère que toutes les entrées non transitives acceptent un argument locatif (Loc N1) qui peut même être parfois obligatoire :

Max s'enfourne dans le tunnel chercher Luc

**Max s'enfourne chercher Luc*

La table 2 a donc comme construction de base N0 V Loc N1 V0-inf W mais également N0 V Loc N1 car l'infinitive peut être supprimée. On a une colonne N0 V V0-inf W qui code la facultativité de l'argument locatif et qui est très souvent codée +, comme c'est le cas dans cet exemple :

Max fonce dans le jardin retrouver ses copains

Max fonce dans le jardin

Max fonce retrouver ses copains

La table 2T a comme construction de base N0 V N1 V0-inf W mais également N0 V N1 car l'infinitive peut être supprimée. On a cette même colonne N0 V V0-inf W qui code ici la possibilité d'omission du N1 et qui est très souvent codée – :

La foule envahit la place manifester sa joie

**La foule envahit manifester sa joie*

2.3. Changement de notation

La construction de base de la table 4 décrite dans [2] (p.170) est Qu P V N1 avec la colonne N0 = : Nnr codée + pour toutes les entrées. Comme l'explique [2], la notation Nnr désigne un substantif « non restreint », c'est-à-dire un substantif d'une nature sémantique quelconque, une complétive ou une infinitive. Le test employé consiste à placer dans la position à étudier les groupes nominaux *Paul* (humain), *ce gâteau* (concret, énumérable, ou non), *la sincérité* (abstrait), *Qu P* et *V-inf W*, les phrases résultantes doivent toutes être acceptables. Nous avons par exemple :

(Paul+Ce gâteau+La sincérité+Qu'il fasse cela+Faire cela) ennue Marie

Nous avons donc supprimé la colonne N0 = : Nnr de la table et l'avons définie comme propriété définitive. Puis, nous avons transformé la construction de base en N0 V N1, puisque Qu P est déjà inclus dans Nnr.

La construction de base de la table 5 décrite dans [2] (p.172) est Qu P V Prép N1 et nous l'avons juste remplacée par N0 V Prép N1 avec la propriété définitive N0 = : Qu P. Ce choix de notation permet de réutiliser les mêmes intitulés dans un maximum de tables et facilite la récupération des informations de distribution des arguments. Ce même procédé a été appliqué à la table 12 qui a N0 V N1 de V1-inf W comme construction de base dans [2] (p.201) et que nous avons remplacé par N0 V N1 de N2 avec la propriété définitive N2 = : V1-inf W.

La construction de base de la table 6 décrite dans [2] (p.177) est N0 V Qu P avec les sept colonnes N1 = : Qu Pind, N1 = : Qu Psubj, N1 = : si P ou si P, N1 = : ce(ci+la), N1 = : Nhum, N1 = : N-hum et N1 = : le fait Qu P codées dans la table. Nous avons gardé N0 V N1 comme construction de base puisque l'argument N1 n'accepte pas toujours la distribution Qu P, et que cela permet de clarifier à quoi font référence ces colonnes. La même méthode a été reproduite :

- les tables 7, 8, 9, 10, 14 et 15 ont N0 V à ce Qu P (respectivement N0 V de Qu P, N0 V Qu P à N2, N0 V Qu P Prép N2, N0 V à ce Qu P Prép N2 et N0 V de ce Qu P Prép N2) comme construction de base dans [2] (p.184, p.187, p.190 et p.193) avec les (ou un extrait des) huit colonnes N1 = : Qu Pind, N1 = : Qu Psubj, N1 = : si P ou si P, N1 = : V0-inf W, N1 = : ce(ci+la), N1 = : Nhum, N1 = : N-hum et N1 = : le fait Qu P codées dans la table ; nous avons modifié leur construction de base en N0 V à N1 pour la table 7, N0 V de N1 pour la table 8, N0 V N1 à N2 pour la table 9 et N0 V N1 Prép N2 pour la table 10.
- les tables 11 et 13 ont N0 V N1 à ce Qu P (respectivement N0 V N1 de ce Qu P) comme construction de base dans [2] (p.197 et p.203) avec les mêmes distributions concernant l'argument N2 codées dans la table ; nous avons remplacé leur construction de base par N0 V N1 à N2 pour la table 11, N0 V N1 de N2 pour la table 13, N0 V à N1 Prép N2 pour la table 14 et N0 V de N1 Prép N2 pour la table 15.
- la table 18 a N0 V Prép N1 Prép N2 (E+pour) Qu Psubj comme construction de base dans [2]

(p.212) avec les mêmes distributions concernant l'argument N3 codées dans la table ; il est vrai que pour la majorité des entrées, la propriété N1 = : Qu Psubj est codée + et que la préposition en position 3 vaut toujours *pour* ou <E> mais une information non valable pour toutes les entrées n'a pas à figurer dans la construction de base, que nous avons donc corrigé par N0 V Prép N1 Prép N2 Prép N3.

2.4. Suppression de classes

Les tables 17 et 19 ont été supprimées, elles avaient il V (E+Prép ce) Qu P (E+Prép N2) (respectivement Qu P V N1 Prép N2) comme construction de base dans [2] (p.211 et 213). La table 19 contenait en effet très peu d'entrées, qui avaient leur place dans d'autres tables de verbes de BGL [3] comme 36DT (*apporter, ôter*), 36SL (*désaccoupler*), 37M4 (*remplir*) et 36R (*servir*), qui ont N0 V N1 Prép N2 comme construction de base (ou N0 V N1 à N2 pour 36R), ce qui correspond bien à la construction initiale. En ce qui concerne les entrées de la table 17, elles sont considérées comme des expressions figées car le sujet *il* est figé. Nous les avons ajouté principalement dans la table d'expressions figées de M. Gross COQ (*il s'agir pour ... de, il apparaître à, etc*) qui a comme construction de base Det0 C0 V Prép1 (Det1 C1+N1) Prép2 N2.

3. Conclusion

Après avoir défini les constructions de base pour chaque classe dans toutes les catégories (verbes, noms prédicatifs, expressions figées et adverbes) [8], ce travail nous a permis de construire un lexique très riche et d'envisager une utilisation de ces données lexicales dans des outils de traitement automatiques [6]. A cette fin, une version textuelle structurée des tables, nommée *LGLex*, a été développée pour les verbes et les noms [1]. La sous-partie de ce lexique qui reproduit les entrées verbales et nominales des tables librement distribuées est elle-même librement distribuée, également sous licence LGPL-LR¹.

Le travail étant plus avancé pour les verbes, qui sont par ailleurs le lexique le plus indispensable dans un analyseur, nous avons pu les intégrer dans un analyseur syntaxique à grande échelle, l'analyseur FRMG [5]. Cette intégration a été possible grâce au travail décrit dans [7], par la conversion des tables au format *Lefff* [4]. L'analyseur syntaxique FRMG couplé à ce lexique a ensuite été évalué sur le corpus de référence de la campagne EASy [7]. Cela valide l'ensemble de l'approche malgré le caractère préliminaire et partiel des résultats.

L'utilisation d'une ressource lexicale la plus riche possible reste donc un moyen efficace pour améliorer la qualité d'un analyseur syntaxique. C'est une des motivations pour poursuivre la construction et l'amélioration des tables du Lexique-Grammaire du français et d'autres langues.

Bibliographie

1. Matthieu Constant et Elsa Tolone. A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables. *Lingue d'Europa e del Mediterraneo, Grammatica comparata*, 1 :79–193, 2010.
2. Maurice Gross. *Méthodes en syntaxe : Régimes des constructions complétives*. Hermann, Paris, France, 1975.
3. Alain Guillet et Christian Leclère. *La structure des phrases simples en français : Les constructions transitives locatives*. Droz, Genève, Suisse, 1992.
4. Benoît Sagot. The lefff, a freely available and large-coverage morphological and syntactic lexicon for french. In *Proceedings of LREC'10*, La Valette, Malte, 2010.
5. François Thomasset et Éric de La Clergerie. Comment obtenir plus des méta-grammaires. In *Actes de la Conférence TALN'05*, Dourdan, France, juin 2005.
6. Elsa Tolone. Les tables du Lexique-Grammaire au format TAL. In *Actes de MajecSTIC 2009*, Avignon, France, 2009. (8 pp.).
7. Elsa Tolone et Benoît Sagot. Using Lexicon-Grammar tables for French verbs in a large-coverage parser. In *Proceedings of the 4th Language and Technology Conference (LTC'09)*, pages 200–204, Poznań, Pologne, 2009.
8. Elsa Tolone, Stavroula Voyatzi, et Christian Leclère. Constructions définitoires des tables du Lexique-Grammaire. In *Actes du 29ème Colloque Lexique et Grammaire (LGC'10)*, Belgrade, Serbie, 2010.